*J. Mol. Biol.* (1995) **254**, 681–691

JMB

# Solution Structure of the GTPase Activating Domain of $\alpha_s$

**D. R. Benjamin[1]*, D. W. Markby[2], H. R. Bourne[2] and I. D. Kuntz[1]**

[1]*Department of Pharmaceutical Chemistry University of California at San Francisco, San Francisco CA 94143-0446, USA*

[2]*Department of Pharmacology University of California at San Francisco, San Francisco CA 94143-0446, USA*

**Corresponding author*

We have used heteronuclear three-dimensional NMR spectroscopy to determine the solution structure of a 141 residue protein containing the GTPase activating domain from the alpha chain of the heterotrimeric G protein $G_s$. The domain contains six α-helices and is stable and structured in solution despite having been excised from the intact $G_s$ protein. The N-terminal ten and C-terminal 11 residues of the protein are unstructured in solution while the core is well determined by the 2483 distance and torsion restraints derived from the NMR spectra. The final ensemble of 14 structures, generated with a hybrid distance geometry/simulated annealing protocol, have an average to-the-mean backbone root-mean-square deviation of 0.39 Å for the core residues 89 to 201. The majority of the structure is remarkably similar to that observed for the cognate domains in crystal structures of the homologous proteins $\alpha_t$ and $\alpha_{i1}$. However, the orientations of the second helix and the subsequent interhelical loops differ markedly among the three proteins. This structural divergence, along with functional studies of chimeric proteins, suggests that this region of the domain interacts with either the downstream effector adenylyl cyclase or with some other intermediary protein.



*Keywords:* NMR; protein structure; simulated annealing; G proteins; adenylyl cyclase

## Introduction

Members of the GTPase superfamily are frequently found among the numerous proteins used by cells for regulating intracellular processes. GTPases are present in organisms as primitive as *Escherichia coli*, and are involved in a range of cellular processes including signal transduction (Birnbaumer, 1990), protein trafficking (Chavrier *et al*., 1990), protein synthesis (Jurnak, 1985) and cytoskeletal rearrangement (Ridley *et al*., 1992). The regulatory GTPases share a common cycle of inter-conversion between an active GTP-bound form and an inactive GDP-bound form, with intrinsic GTP hydrolysis and GDP release driving the cycle (Bourne *et al*., 1991). The rates of both GTP hydrolysis and GDP release are regulated by a variety of other proteins in the cell, allowing precise control over the proportions of active and inactive intracellular GTPases. New members of the superfamily are still being discovered (Kehlenbach *et al*., 1994) but it is already clear that the use of GTP hydrolysis in the regulation of cellular events is an ancient mechanism.

The heterotrimeric G proteins are a family of GTPases that effect signal transduction. They couple activation of serpentine receptors (membrane-bound proteins with seven membrane-spanning segments) to the regulation of a variety of downstream effectors. These proteins are composed of a GTP binding alpha chain and a tightly associated complex of beta and gamma chains. Sequence comparisons suggest that $G_\alpha$ subunits contain two discrete domains: a core GTP-binding domain identical in topology with the α/β fold of p21 ras and EF-Tu plus a second ~120 residue domain, unique to heterotrimeric members of the superfamily, inserted within a loop of the GTPase core.

The structure of the inserted domain was first revealed in the crystal structures of alpha subunits from the G proteins transducin ($\alpha_t$; Noel *et al*., 1993) and $G_{i1}$ ($\alpha_{i1}$; Coleman *et al*., 1994). The domain consists of five short alpha helices arranged about a long central helix. The backbone structures of the inserted domains in $\alpha_t$ and $\alpha_{i1}$ are superimposable

Abbreviations used: RMSD, root-mean-square deviation; AOP, angular order parameters; nOe, nuclear Overhauser enhancement.

0022–2836/95/490681–11 $12.00/0

with the exception of the C terminus of the second helix and the loop between the second and third helices. Bound guanine nucleotide is buried in a cleft between the domains with direct protein-ligand interactions confined largely to the $\alpha/\beta$ domain.

Despite the availability of detailed structural information along with a variety of biochemical and molecular genetic data for different G proteins, the function of the inserted domain remains largely unknown. Several hypotheses have been advanced: the observation that a guanine nucleotide lies buried in the interdomain cleft prompted the suggestion that the domain serves as a lid on the nucleotide binding site and that receptors catalyze nucleotide exchange by promoting ''en bloc'' movement of the domain to allow GDP release (Noel *et al*., 1993). The limited sequence similarity shared among the inserted domains in different G proteins led to the proposal that they served as interaction domains for downstream effectors (Masters *et al*., 1986), although more recent molecular genetic studies localize interaction determinants to the core GTPase domain (Conklin & Bourne, 1993). The hypothesis that the domain serves as an intrinsic GTPase activating protein (or GAP; Freissmuth & Gilman, 1989; Landis *et al*., 1989) takes into account the relatively rapid GTP hydrolysis rate of $G_\alpha$ subunits compared with small GTPases (e.g. p21 Ras). The GTPase activity of small GTPases is stimulated by exogenous GAP proteins, which bind the GTPase core at the location of the insert. This analogy between the inserted domain and GAP is supported by the demonstration that an isolated domain stimulates GTP hydrolysis in a separately expressed $\alpha_s$ GTPase core (Markby *et al*., 1993). We therefore refer to this inserted domain as the GA (Gtpase-Accelerating) domain.

The recent advent of multidimensional heteronuclear NMR methods has enabled the determination of solution structures of proteins with a molecular mass of up to 25 kDa (Wagner, 1993). These methods provide an independent means to obtain tertiary structure and are the only route to high-resolution structures in solution. Additionally, NMR spectra can measure the time-scales for various dynamic processes in proteins (Bruschweiler & Case, 1994). We have used NMR spectroscopy to determine the solution structure of the GA domain of rat $\alpha_s$, providing the first tertiary structural information for this G protein and allowing us to investigate the structural similarities and differences in the GA domains of $\alpha_s$, $\alpha_t$ and $\alpha_{i1}$.

## Results

### Calculations

A total of 2297 interproton distance restraints, and 86 coupling constants were obtained from NMR spectra of doubly labeled ($^{15}N$, $^{13}C$) $\alpha_s$ GA domain: 313 long-range, 424 medium-range, 583 sequential and 977 intraresidue nOes comprise the distance restraints. Stereospecific assignments for 39% of all prochiral protons and methyl groups were made as described in Materials and Methods. Identification of amide protons protected from chemical exchange with solvent yielded 100 hydrogen bonding restraints (50 HN-O and 50 N-O). Distance geometry calculations incorporating all restraints generated 100 structures; the ten structures with the lowest overall constraint violations were then used to initiate two sequential rounds of simulated annealing, each yielding 100 different structures. An ensemble of 14 structures selected from this group constitutes the final set of structures, and has been submitted to the Brookhaven Protein Data Bank and are available directly from the authors on request until they have been processed and released.

### Structure description

Throughout this work we number residues by their positions in the long splice variant of $\alpha_s$. In this scheme, the protein we studied encompasses residues 73 to 214. The precision of the final ensemble is evident in Figure 1, which shows a stereo representation of the overlay of the backbone atoms from the final ensemble of 14 simulated annealing structures and the energy-minimized average structure. The local RMSD values and number of nOes per residue are graphed in Figure 2. The number of long-range nOes per residue exhibits a periodicity arising from the protein's helical nature; the helices have approximately every third residue pointing into the interior of the protein. The angular order parameters (AOP; Hyberts *et al*., 1992) for the torsion angles $\phi$ and $\psi$ are shown in Figure 3. These Figures show that the domain backbone is quite well defined by the NMR data with the exception of the flexible termini and three interhelical loops. The average to-the-mean RMSD for all heavy-atoms is 1.33 Å, and for backbone atoms ($C^\alpha$, C', N) is 1.15 Å; for the core region of residues 89 to 201 the corresponding figures are 0.78 Å and 0.39 Å. Many side-chains in the protein are also well ordered, as indicated by the structures shown in Figure 4. Side-chains in the protein interior are particularly well ordered; solvent-accessible chains typically have fewer nOes and adopt a variety of rotameric states, indicating that the protein samples conformational space in these regions. The side-chain order is partly attributable to the availability of stereospecific assignments, which have been shown to improve the precision of NMR structures (Garrett *et al*., 1994).

The structural statistics presented in Table 1 indicate that the ensemble is also energetically reasonable. There is no nOe violation above 0.4 Å, the structures show good covalent geometry in terms of bond lengths and angles, and the negative van der Waals energy indicates the absence of non-bonded overlap. A Ramachandran plot of all residues in the structures is shown in Figure 5(a); the same plot for residues having $\phi,\psi$ AOPs > 0.9 is in Figure 5(b). The helical content of the protein is

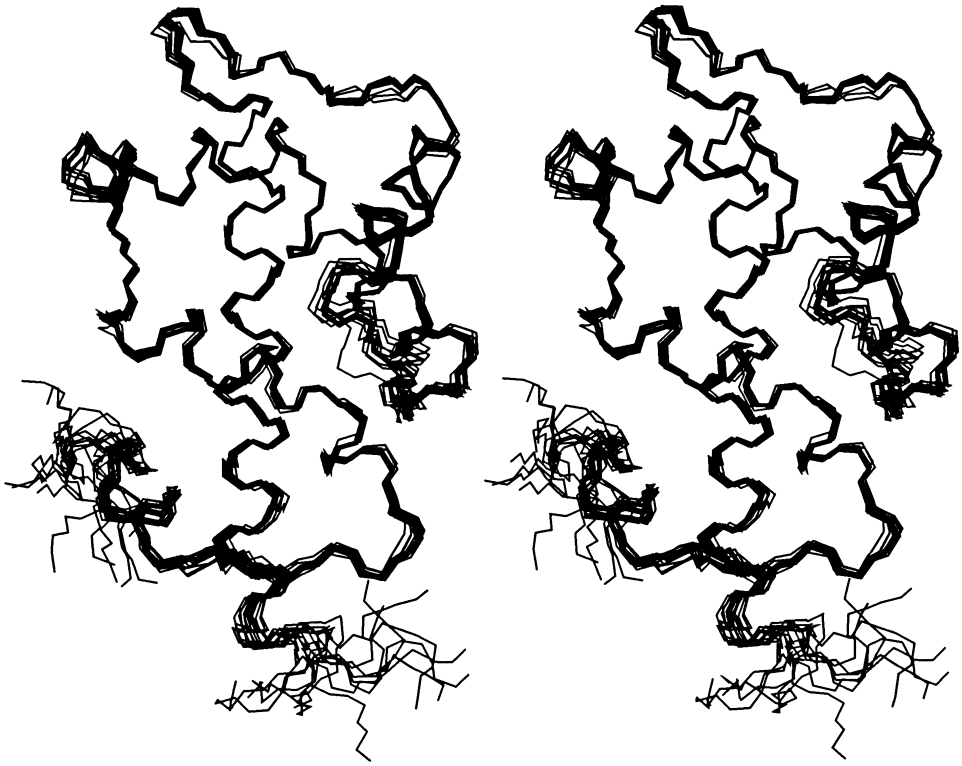

**Figure 1.** Stereo superposition of N, $C^{\alpha}$, C', O atoms from final ensemble of 14 structures and minimized mean structure for the $\alpha_s$ GA domain.

evident in the preponderance of residues with $\phi,\psi$ angles of –60°, –30°. Almost all $\phi$ and $\psi$ angles in the core region fall within energetically favorable bounds, while the less ordered terminal residues adopt unfavorable conformations, a pattern observed in many NMR structures. This deviation presumably arises from unrestrained residues sampling high-energy conformations during the molecular dynamics simulations and getting trapped in local minima in the subsequent energy minimization.

The overall fold adopted by the GA domain of $G_s$ chains has not been identified in any other protein studied to date (Noel *et al*., 1993). Six alpha helices

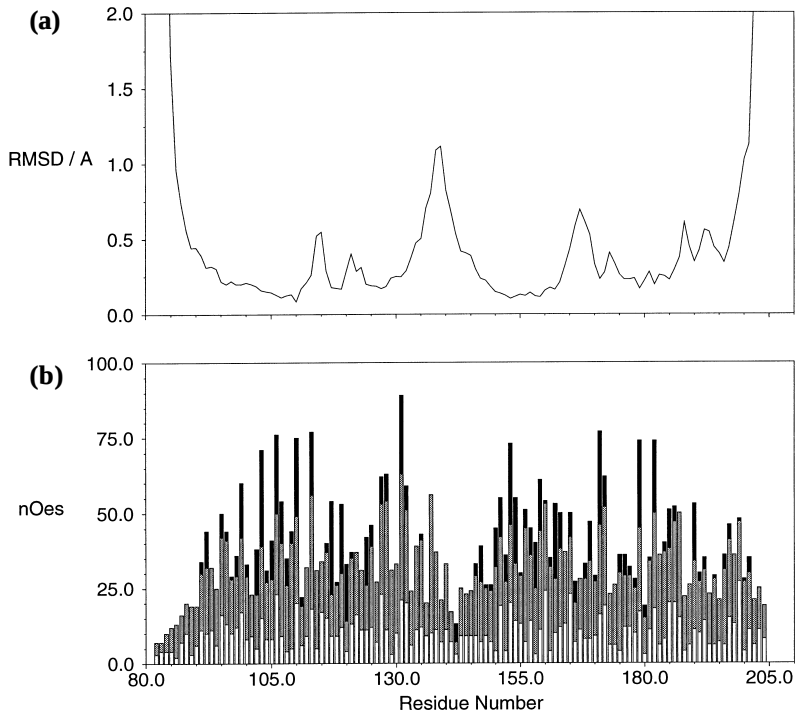


**Figure 2.** Per-residue structural statistics in the ensemble of 14 structures. (a) Local backbone ($C^{\alpha}$, C', N) RMSD in atomic positions. Residues 81 to 83 and 202 to 204 are plotted offscale; RMSD values for these residues ranged from 3.0 to 6.5 Å. (b) Distribution of nOes per residue. Bars represent intraresidue (black bar), medium-range (hatched bar) and long-range (open bar); the height of a bar represents the sum of all nOes. Medium and long-range nOes appear twice in this Figure, at both source and destination residue.

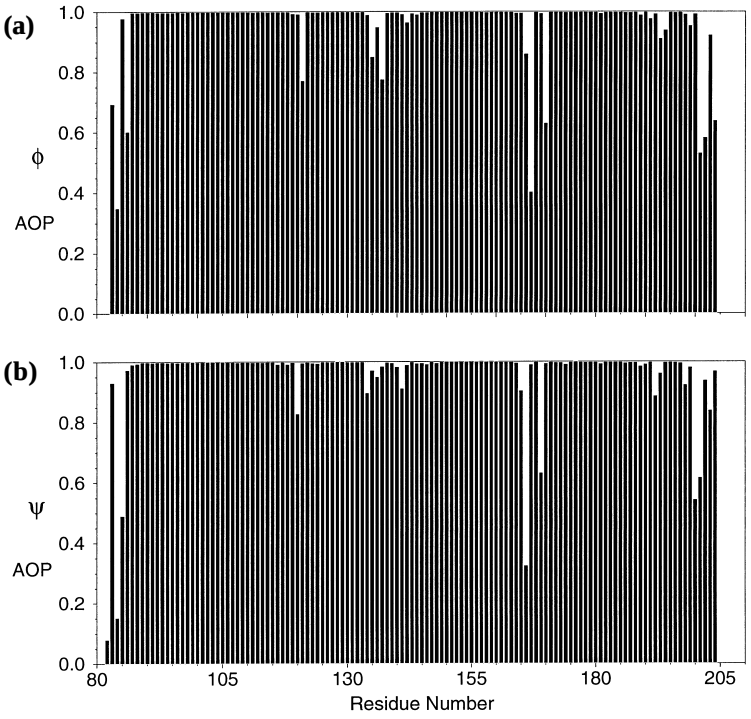


**Figure 3.** Per-residue angular order parameters. (a) ϕ torsion angle; (b) ψ torsion angle.

(A to F) are the only regular secondary structures present in the $\alpha_s$ GA domain. The A helix, extending from residues 86 to 110, is the backbone of the domain around which the other five helices are arranged. The N terminus of this helix is ragged and is clearly not persistently helical in the isolated domain. It also has elements of $3_{10}$ helix in some members of the ensemble, consistent with the proposal that $3_{10}$ helices appear in partially unfolded helical peptides (Millhauser, 1995). The A helix is well defined, and much of it is protected from solvent by the other helices. At the end of the A helix is a ten residue loop that includes a *cis* Proline (P115), the site of a two amino acid insertion in $\alpha_s$ relative to other alpha subunits. Curiously, three residues in this ten residue loop (Leu119 to Asn121) have exchange protected amide protons with no obvious hydrogen-bonding partners. Ala120 is almost entirely excluded from solvent, which may explain its protection, but the other two amide protons are solvent-accessible.

The B helix, extending from residues 122 to 134, is one of two helices in the GA domain to start with a proline residue. Members of the ensemble show some small variation in the orientation of this helix with respect to the rest of the protein. The B helix is followed by a loop representing the most poorly defined area of the core. The loop, which stretches from residues 135 to 141, has several residues with HN-$H^{\alpha}$ coupling constants of 6 to 8 Hz, indicative of torsional averaging about the N-$C^{\alpha}$ bond. There is also virtually no long-range nOe in this region (Figure 2). The loop and the N terminus of the subsequent helix contain five aromatic residues in a

**Figure 4.** Stereo superposition of ensemble of 14 $\alpha_s$ GA domain structures showing all heavy-atoms for residues 96 to 133. Interior residues are well ordered, while solvent-exposed side-chains show conformational variation.

**Table 1.** Structural statistics

| RMSD from experimental distance restraints (Å) | |
|---|---|
| All | 0.044 ± 0.001 |
| Intra | 0.053 ± 0.001 |
| Sequential ($\|i-j\|=1$) | 0.032 ± 0.002 |
| Short ($1<\|i-j\|<5$) | 0.039 ± 0.002 |
| Long ($\|i-j\|>5$) | 0.040 ± 0.001 |
| H-bonds | 0.031 ± 0.001 |
| Deviations from idealized geometry | |
| Bonds (Å) | 0.009 ± 0.0003 |
| Angles (deg) | 2.93 ± 0.37 |
| AMBER energies (kcal/mol) | |
| $E_{total}$ | –346.9 ± 10.0 |
| $E_{AMBER}$ | –453.4 ± 11.5 |
| $E_{constraint}$ | 106.5 ± 3.8 |
| $E_{angle}$ | 238.8 ± 4.7 |
| $E_{vdW}$ | –727.2 ± 8.8 |
| $E_{bond}$ | 32.3 ± 0.6 |
| $E_{dihed}$ | 203.0 ± 5.6 |

ten residue stretch. The aromatic ring protons have degenerate resonant frequencies, also suggesting averaging of the environment and preventing unambiguous assignment of a number of potentially restraining nOes. All of these observations are consistent with dynamic averaging of the interhelical loop. Surprisingly, the amide proton of F142 has a resonant frequency of 5.10 ppm; this large shift from the random coil value of 8.18 ppm suggests a ring current effect arising from a fixed orientation relative to an aromatic ring. This shift is not consistent with structural averaging and suggests that there may be some stable substructures within this conformationally flexible loop.

Proline 144 marks the N terminus of the C helix, which runs nearly parallel with helix A and continues through to Asp156. An unusual turn with sequential $(i, i+3)$ and $(i, i+5)$ hydrogen bonds links the C and D helices; chemical exchange protection of amide protons in this turn indicate that the hydrogen bonds are persistent. Helix D includes Cys162, which is entirely shielded from solvent. The sulfhydryl proton of this residue is protected from chemical exchange and has a chemical shift of 3.14 ppm. nOes originating from this proton were unassigned until late in the refinement stage; these nOes were asymmetrical (i.e. there was no ''mirror'' nOe to a proton at 3.14 ppm in either the $^{15}N$ or $^{13}C$ NOESY spectra) and were absent in samples dissolved in $^2H_2O$ for a week. After refinement, Cys162 $H^{\gamma}$ was the only exchangeable proton consistent with the unassigned nOes.

A ten residue loop links helix D to helix E, the most distorted helix. Helix E is nearly orthogonal to helix A and has a pronounced kink at Lys181 in order to maintain side-chain contacts along its length. The distortion is so severe that it complicates precise delineation of the helical endpoints. The amide protons in the middle of this helix are not protected from chemical exchange, indicating that these hydrogen bonds are short-lived.

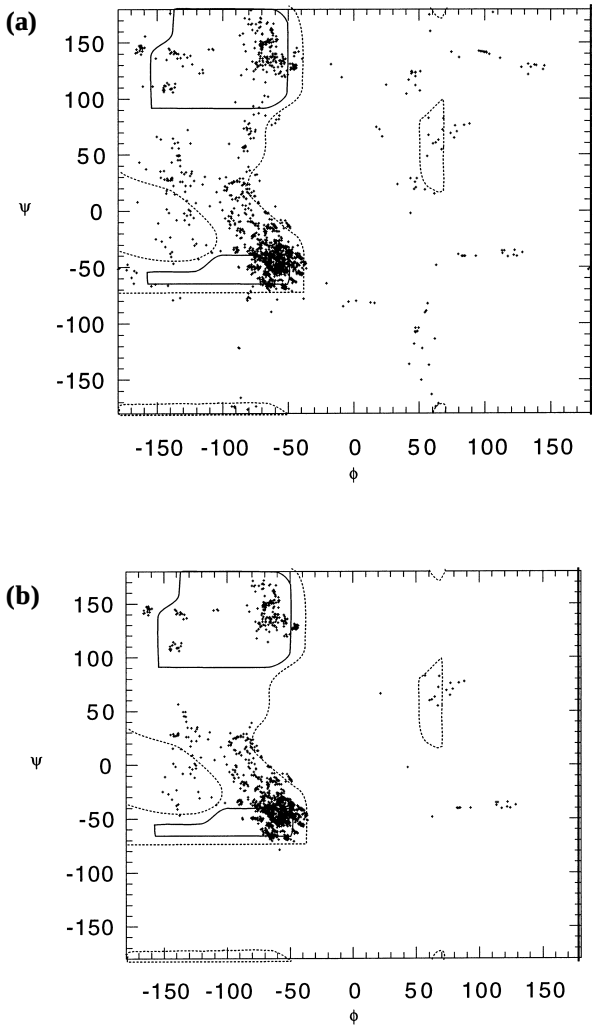


**Figure 5.** Ramachandran plots for (a) all residues, (b) residues having $\phi,\psi$ angular order parameters >0.9.

Helix F is the last helix in the isolated domain and shows considerable fraying. It exhibits no amide proton protection and the subsequent C-terminal residues are entirely unstructured. The amide proton of R201, at the end of helix F, has a resonant frequency (9.90 ppm) that is the third most downfield in the molecule, suggesting that the backbone has a defined local structure. However, examination of the ensemble of structures shows that the F helix adopts a number of conformations relative to the rest of the protein. There is also a paucity of long-range nOes involving helix F. Thus, this helix appears to be perturbed by end effects that are probably absent in the intact protein.

## Comparison with $\alpha_i$ and $\alpha_t$

Overlays of the minimized average NMR structure of the $\alpha_s$GA domain along with the GA domains from the proteins $\alpha_t$ and $\alpha_{i1}$ are shown in Figure 6. The similarity among the proteins is striking, especially considering that they share only about 20% amino acid identity, that the $\alpha_t$ and $\alpha_{i1}$

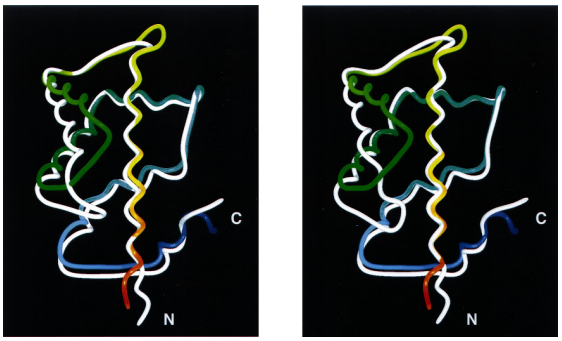


**Figure 6.** Superposition of residues 8 to 120 of (shown in color) with (a) residues 63 to 174 of (shown in white) and (b) residues 69 to 176 of $\alpha_i$ (shown in white). The superposition was calculated using residues 8 to 31, 36 to 52 and 64 to 120 of $\alpha_s$, residues 63 to 86, 89 to 105, and 118 to 174 of $\alpha_t$ and residues 69 to 90, 94 to 101, and 130 to 176 of $\alpha_i$. Carbon alpha tubes are shown, with as GA domain ranging from red at the N terminus to blue at the C terminus.

structures were obtained by X-ray crystallography, and that the $\alpha_t$ and $\alpha_{i1}$ structures are in the context of the intact protein while the $\alpha_s$ structure is of the isolated domain. It is relevant also that the AMBER (Weiner *et al*., 1986) force-field was used to refine the NMR structure, while the XPLOR (Brunger *et al*., 1987) force-field was used for the crystallographic refinement. Thus, the structural similarity derives from the data and not from force-field bias in the refining protocol.

Excluding the B helix, the B/C loop, and residues 111 to 117 (containing the two residue insertion in $\alpha_s$), the backbone ($C^\alpha$, C', N) RMSD between the average $\alpha_s$ and the $\alpha_t$ GA domain is 1.20 Å. The overlap of $\alpha_s$ GA domain with $\alpha_{i1}$ is slightly better, giving an RMSD of 1.10 Å. It is noteworthy that there are three G protein/nucleotide complexes in the asymmetric unit of the GTPγS $\alpha_t$ crystal structure (Noel *et al*., 1993); the average backbone RMSD between GA domains from these three complexes is 0.43 Å. Additionally, the backbone RMSD between the GA domains in the GDP (Lambright *et al*., 1994) and GTPγS $\alpha_t$ crystal structures is 0.79 Å, and between the GTPγS forms of $\alpha_t$ and $\alpha_{i1}$ (Coleman *et al*., 1994) is 0.87 Å. Hence, in the region of greatest similarity the NMR structure of $\alpha_s$ GA domain has a backbone that is virtually indistinguishable from that of the other alpha subunits.

The region excluded from the backbone comparisons above (residues 125 to 144) was determined by plotting the RMSD between the proteins as a function of increasing gap length, with a gap centered at residue 139. The inflection in this curve was used to define the border between the similar and distinct regions of the proteins. Inspection of Figure 6 shows that the orientation of helix B differs among the proteins. Helix B is of identical length in $\alpha_s$ and $\alpha_t$; it is extended by an additional turn of $3_{10}$ helix in $\alpha_{i1}$. The B/C loop, which is altered both in structure and in length among the three proteins, is particularly interesting in that it is the only region where the structure of $\alpha_t$ GA domain differs significantly from that of $\alpha_{i1}$. It is also the least defined region in the core of the NMR structure and the region with the highest *B* factors in the core of the $\alpha_t$ crystal structure. Although the NMR structures may overestimate the flexibility of this region (see above), the loops are clearly different among the three proteins and are apparently more mobile than the rest of the GA domain.

One concern with studying the tertiary structure of a domain excised from the intact protein is that the absence of intraprotein contacts may result in artifactual structural rearrangements in the isolated domain. In the case of the GA domain this fear was not realized. The GA domains of $\alpha_t$ and $\alpha_{i1}$ are largely autonomous and have few contacts with the GTPase cores. The model in Figure 6 is oriented such that the alpha chain GTPase core is on the right-hand side; helix D and the D/E loop are the only regions in the GA domain that make contact with the GTPase core. These regions also exhibit the highest sequence conservation among various GA domains, consistent with a role of engaging in interdomain contact with the more conserved GTPase core. Comparison of the three separately determined structures shows that in these areas the $\alpha_s$ backbone is superimposable with those of $\alpha_t$ and $\alpha_{i1}$. The fact that residues close to the GTPase core are unperturbed by its removal indicates that the structural differences we observe on the other side of the isolated domain are genuine, and not artifacts of our model system.

### Sequence alignment

All G protein alpha subunits cloned to date have a GA domain; however, the sequence similarity among the various domains is low and complicates sequence alignment. The availability of tertiary structures for three GA domains allows us to align the sequences based on conservation of structural elements; this alignment is presented in Figure 7. The differences relative to sequence-based alignments are that (1) Pro115-Pro116 in $\alpha_s$ represents a two residue insertion and (2) there is a one residue deletion in $\alpha_s$ in the region from residue 135 to residue 144. The alignment in this region is approximate in that the exact location of the one residue deletion is not obvious, as the local structures of the three proteins differ greatly. However, maintaining the prior and subsequent helices in register requires a single residue deletion somewhere in this loop.

## Discussion

### Structural implications

With 124 of 145 residues present in the $\alpha_s$ GA domain being structured in solution, the domain represents one of the largest protein structures determined by NMR to date. The low RMSD values

and high angular order parameters for the ensemble show that the domain core is well defined in solution with the exception of the loop between helices B and C. In small GTPases (e.g. p21 Ras, EF-Tu) the guanine nucleotide is partially accessible to solvent and apparently free to diffuse away after hydrolysis (although these proteins exhibit a high affinity for guanine nucleotide). In G proteins the nucleotide is entirely shielded from solvent and there is no obvious route for nucleotide release. The NMR structure shows that the GA domain is autonomous and supports the hypothesis (Noel *et al*., 1993) that the rigid GA domain moves en bloc in order to allow GDP release. The stability of a GA domain in the absence of about 220 residues from the alpha chain is consistent with the observation of relatively few interdomain contacts in crystal structures of intact alpha chains, and indicates that the guanine nucleotide is the ''glue'' positioning the two domains together.

Earlier work confirmed the hypothesis that the GA domain serves as a GTPase accelerating domain when added to a separately expressed GTPase core (Markby *et al*., 1993). Residue R201 in the GA domain has been implicated as a catalytic residue in both biochemical and crystallographic studies. Although the backbone atoms of this residue are reasonably well defined in the excised domain, the side-chain is essentially unconstrained. We are therefore unable to comment on any structural role played by this residue. Additionally, the GA domain is unable to complement adenylyl cyclase stimulation by the GTPase core unless residues 204 to 213 are included. These residues form beta strand 2 in the intact protein and are entirely unstructured in the isolated domain. Presumably these residues are ordered in the complex, either displacing the corresponding residues in the GTPase core or extending the beta sheet.

It has been shown recently that a rare combined syndrome, pseudohypoparathyroidism-testotoxicosis, can be caused by an alanine to serine site mutation in the GTPase domain of $\alpha_s$ (Iiri *et al*., 1994). Biochemical investigation of this mutant reveals that it is constitutively activated due to an accelerated intrinsic GDP off rate. The mutant is also thermolabile and undergoes a rapid denaturation at physiological temperature. The stability of the isolated GA domain suggests that this rapid denaturation occurs within the GTPase core and not the GA domain.

## Functional implications

The GTPase core in the alpha chain of $G_s$ is present in all GTPase superfamily members; the GA domain is unique to G proteins. In small GTPases (e.g. p21 Ras), the short loop into which the GA domain is inserted is known as the ''effector loop'', based on the observation that site-directed mutants in this area prevented constitutively activated proteins from transforming cells without altering nucleotide binding (Sigal *et al*., 1986). The GA domains of various alpha chain subfamilies have very limited sequence similarity, suggesting that the domains contain type-specific sequences; however, studies of chimeric proteins indicate that the effector interaction regions are within the GTPase core of the alpha chain (Berlot & Bourne, 1992) and it has been assumed that the GA domain is not involved in effector coupling.

More recent evidence suggested that the GA domain of $\alpha_s$ may play a role in effector coupling. Antonelli *et al*. (1994) demonstrated that the *Xenopus laevis* $\alpha_s$ cannot stimulate the human adenylyl cyclase unless a 70 residue segment from the human $\alpha_s$ GA domain is substituted into the *Xenopus* sequence. This substitution changes only 19 residues, eight of which are within the B helix and B/C loop. Figure 8 shows two views of the GA domain, with residues differing between the human and *Xenopus* sequences colored red and shared residues colored white. The Figure shows that seven of eight B/C substitutions, along with four others in the GA domain, are clustered along one face of the polypeptide. This face is on the same

```
          81                                                  130
gbt1_bovin  DGYSLEECLE FIAIIYGNTL QSILAIVRAM TTLN..IQYG DSARQDDARK
  gbi1_rat  AGYSEEECKQ YKAVVYSNTI QSIIAIIRAM GRLK..IDFG DAARADDARQ
gbas_mouse  RSNSDGEKAT KVQDIKNNLK EAIETIVAAM SNLVPPVELA NPENQFRVDY
gbas_xenl   NGFNAEEKKT KVQDIKNNIK EAIETIVTAM GNLSPPVELV NPENQFRIDY

          131                                                 180
gbt1_bovin  LMHMAMDTIE EGTPKEMSDI IQRLWKDSGI QACFDRASEY QLNDSAGYYL
  gbi1_rat  ALFVLMGAAE EGFTAELAGV IKRLWKDSGV QACFNRSREY QLNDSAAYYL
gbas_mouse  ILSVMFNVP. NFDPPEFYEH AKALWEDEGV RACYERSNEY QLIDCAQYFL
gbas_xenl   ILNLPNYKD. FEFSPEFYEH TKTLWQDEGV RACYERSNEY QLIDCAQYFL

          181                              209
gbt1_bovin  SDLERYLVTP GVPTEQDVLR SRVKTTGIIE
  gbi1_rat  NDLDRYIAQP NIPTQQDVLR TRVKTTGIVE
gbas_mouse  DKIDVYIKQA DVPSDQDLLR CRVLTSGIFE
gbas_xenl   DKIDIVKQND YTPSDQDLLR CRVLTSGIFE
```

**Figure 7.** Sequence alignment of GA domains of $\alpha_s$, $\alpha_t$, and $\alpha_{i1}$ based on conservation of tertiary structure, along with the sequence for the *Xenopus laevis* $\alpha_s$ GA domain.

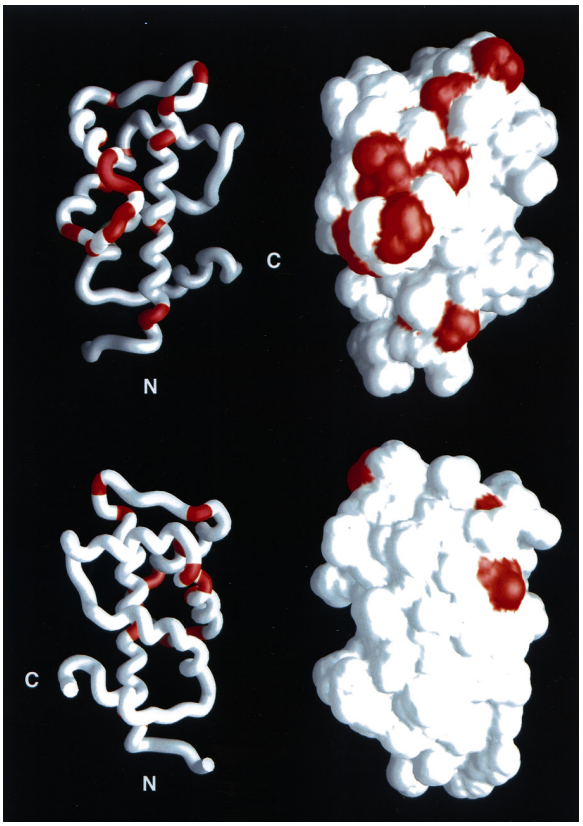


**Figure 8.** Amino acid substitutions in *Xenopus laevis* $\alpha_s$ relative to human $\alpha_s$. Residues that are both present in the GA domain and changed in the *Xenopus* sequence are colored red. Carbon alpha tubes are on the left and solvent-accessible surfaces on the right. The top and bottom Figures differ by a 180° rotation about the *y* axis.

side of the intact molecule as switch II, a region associated with alpha chain activation (Lambright *et al*., 1994). Thus, the location of these substitutions is consistent with a role in effector coupling. Another supporting observation for this role comes from a molecular genetic study (Wilson & Bourne, 1995) in which it was noted that changing four residues in the B/C loop and the N terminus of helix C resulted in the expression of a protein with no signaling activity even after treatment with cholera toxin.

The conformational flexibility of the B/C loop is in keeping with that observed in active sites or binding loops of several other protein structures determined by NMR methods (Martin *et al*., 1995; Fushman *et al*., 1995; Hyberts *et al*., 1992). Given the functional observations described above, and the structural differences in the B/C loop between $\alpha_s$ and other alpha subunits, we propose that the B/C loop (and possibly the B helix) is involved in $\alpha_s$ coupling to adenylyl cyclase, either through direct contacts or through an intermediary protein (e.g. the beta gamma subunit of the heterotrimer or a second alpha subunit). A possible mechanism for this proposal is suggested by comparison of the GDP and GTP$\gamma$S forms of $\alpha_{i1}$ (S. Sprang, personal communication). GDP-form crystals of $\alpha_{i1}$ exhibit a dimer of alpha subunits in the asymmetric unit. In the dimers, the N and C termini of one subunit are in direct contact with the C terminus of the B helix, the B/C loop and the N terminus of the C helix. This contact unwinds the extra turn of $3_{10}$ helix observed at the end of the B helix in the GTP$\gamma$S form, and moves the B/C loop into a conformation similar to that observed in $\alpha_t$. This observation suggests that the GA domain may be involved in oligomerization of alpha subunits.

Our proposal must be considered in light of studies showing that $\alpha_s/\alpha_{i1}$ chimeras containing $\alpha_{i1}$ GA domains are apparently competent to stimulate adenylyl cyclase, albeit with a higher basal level of activity (Berlot & Bourne, 1992), and that the *Drosophila melanogaster* $\alpha_s$ subunit (Quan *et al*., 1991) and the $\alpha_{olf}$ subunit (Jones *et al*., 1990) show activity in assays with human adenylyl cyclase despite having almost no sequence similarity to human $\alpha_s$ in the B/C loop. These observations are somewhat inconsistent with our proposal for the role of the GA domain. Interpretation of all of the functional studies is complicated by the fact that adenylyl cyclase stimulation is influenced by a number of alpha chain properties, including the intrinsic rates of GTP hydrolysis and GDP release, protein expression levels, affinity for beta/gamma subunits, and affinity for adenylyl cyclase. Resolution of the issues raised by the molecular genetic experiments will require the concerted application of enzymology, protein biochemistry and structural biology.

## Materials and Methods

### NMR spectroscopy

Complete assignments and secondary structure of $\alpha_s$ GA domain have been reported elsewhere (Benjamin *et al*., 1995). Samples (2.5 mM) of uniformly $^{15}N$ and $^{13}C$, $^{15}N$-labeled recombinant $\alpha_s$ GA domain (in 10 mM sodium phosphate (pH 6.0), 10 mM dithiothreitol, 0.005% (w/v) sodium azide) were prepared as described by Benjamin *et al.* (1995). NMR spectra were acquired at 303 K on a Varian Unity-plus 600 MHz spectrometer equipped with a triple resonance probe and actively shielded field gradients. Distance constraints were obtained from analysis of the 3D $^{13}C$ HSQC-NOESY (Muhandiram *et al*., 1993) and $^{15}N$ HSQC NOESY (50, 100 and 150 ms mix times). An HNHA-J spectrum (Vuister & Bax, 1993) was acquired for measurement of $\phi$ coupling constants. Qualitative C'-$H^{\beta}$ coupling constants were obtained from an HNCOHB experiment (Grzesiek *et al*., 1993) and amide proton exchange protection was monitored *via* an $^{15}N$ HSQC experiment using a semi-constant time evolution period (Davis, 1995). States-TPPI phase cycling was used for frequency discrimination in indirect dimensions (Marion *et al*., 1989). Heteronuclear dimensions incorporated 180° linear phase ramps to allow unambiguous identification of aliased peaks arising from the small sweep widths employed (Bax *et al*., 1991). HN detected spectra were acquired using the sensitivity-enhanced gradient selection scheme of Kay *et al*. (1992) and included a water flip-back pulse to minimize water saturation (Grzesiek & Bax, 1993). Data were processed with the nmrPipe suite of programs (F. Delaglio, unpublished) and analyzed with the in-house program

Sparky (D. Kneller, unpublished). All spectra were extended in heteronuclear dimensions by forward-backward linear prediction (Zhu & Bax, 1992) prior to apodization and zero filling.

### Constraints

Distance constraints were obtained from the 100 ms 3D NOESY spectra. Peaks were classified as strong, medium or weak using contour levels for calibration; the corresponding distance restraints were 1.8 to 2.8 Å, 1.8 to 3.5 Å and 1.8 to 5.0 Å. Upper distance restraints involving methylene, methyl and aromatic protons that had not been stereospecifically assigned were augmented with appropriate pseudoatom corrections (Wüthrich, 1986). Quantitative coupling constants for the HN-$H^{\alpha}$ torsion angle of 114 residues were obtained from the HNHA-J spectrum using published procedures (Vuister & Bax, 1993). Coupling constants between 5 and 8 Hz, presumably arising from torsional averaging, were excluded from the refinement. The remaining 86 constants were converted into loose torsional restraints for distance geometry calculations only, with coupling constants <5 Hz giving $-30° < \phi < -90°$ and constants >8 Hz giving $-90° < \phi < -150°$; simulated annealing runs were refined directly against the coupling constants (Garrett *et al*., 1994). 56 amide protons exhibit protection from chemical exchange with $^2H_2O$. Of these, 50 were assigned hydrogen-bonding partners on the basis of preliminary distance geometry calculations, resulting in 100 hydrogen-bonding restraints (upper limits of 2.3 Å for HN-O and 3.3 Å for N-O).

### Stereospecific assignments

The resonance assignments presented earlier (Benjamin *et al*., 1995) remain unchanged. Stereospecific assignments of 51 resolved $H^{\beta}$ protons were obtained by examination of the 50 ms 3D $^{15}N$ and $^{13}C$ NOESY spectra and the HNCOHB spectrum. The distance geometry program DIANA (Güntert *et al*., 1991) was used to generate a preliminary set of 50 structures from the nOe and torsional restraints described above, using pseudo atom corrections for all restraints involving non-$H^{\beta}$ prochiral protons. No hydrogen-bonding restraint was included in these preliminary calculations. The 15 structures with the lowest DIANA scores were energy minimized with the SANDER module of AMBER (Weiner *et al*., 1986) using an all-atom force-field and 5 kcal/mol force constants for the distance restraints. The resulting structures were examined graphically; stereospecific assignments were made for those spectroscopically resolved protons and methyl groups where at least 12 of the structures had identical rotamers and where a distinguishing pattern of medium or long-range nOes predicted from the structures was evident in the data. In all, 39% of the prochiral groups were stereospecifically assigned. The preliminary structures were also used to identify hydrogen-bonding partners for exchange-protected amide protons and to assign ambiguous nOes.

### Structure calculations

All calculations were performed using SGI and HP workstations. DIANA was used for distance geometry calculations and the AMBER suite of programs for energy minimization and molecular dynamics. The AMBER all-atom force-field was employed for minimization; modifications for simulated annealing included reducing the charges on a charged side-chain to 0.2 and increasing proton masses to 10 Da. Chirality and planar amide bond restraints were included in annealing runs. Residues 73 to 81 (containing a monoclonal antibody epitope used for developing the purification train) and 204 to 213 were found to be unstructured in solution and were not included in any calculations. A family of 100 structures was generated with distance geometry using the REDAC strategy (Güntert & Wüthrich, 1991). The 14 structures with the lowest constraint violations were subjected to 100 iterations of steepest descent and 400 iterations of conjugate gradient energy minimization using an nOe constraint force constant of 5.0 kcal/mol. Each structure was then used for ten different simulated annealing runs, using a different seed for the random velocity generator each time and nOe restraints raised from 5 to 30 kcal/mol during the first 2 ps. The annealing protocol consisted of 1 ps of dynamics where the temperature was raised from 10 K to 1200 K, 2 ps of dynamics at 1200 K, and a 12 ps annealing period ending at 0 K. The temperature coupling constants were 0.2 for the first 3 ps, 2.0 for the next 8 ps, 1.0 for 2 ps, 0.5 for the next 1 ps, and 0.05 for the final picosecond of dynamics: 500 steps of energy minimization with restraints ended the first round of calculations. The ten structures with the lowest overall energies (force-field + restraint) were then subjected to a second round of annealing and minimization to improve convergence (Moore *et al*., 1991); a force constant of 30 kcal/mol was used for nOe restraints throughout this second round, yielding 100 structures. The final ensemble consists of the 14 structures with the lowest overall energies and having no nOe violations above 0.4 Å and no distortions of bond lengths greater than 0.02 Å or bond angles greater than 5°.

### Structural analysis

MIDASPlus (Ferrin *et al*., 1988) was used for all graphical analysis and the generation of protein structure figures. Figure 8 was generated with the molecular graphics program GRASP (Nicholls *et al*., 1991). The protein structure analysis program DISTAN (D. Deerfield, unpublished) was used to calculate torsion angles, deviations in bond lengths and angles, and solvent accessibility. The average structure was obtained with the program CARNAL by overlaying all 14 structures with a least-squares fitting of backbone atoms for residues 81 to 201. The atomic coordinates were averaged, and the resulting structure subjected to 1000 steps of energy minimization with constraints.

## Acknowledgements

We thank Drs Don Kneller, David Deerfield and Frank Delaglio for providing software, and Dr Vladimir Basus for assistance with hardware. We acknowledge support from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081). We also acknowledge Dr Lewis E. Kay, Dr David A. Case and Tom Cheatham for helpful discussions. D.R.B. was supported by an NSF Graduate Fellowship and a U.C. Chancellor's Graduate Research Fellowship.

## References

Antonelli, M., Birnbaumer, L., Allende, J. E. & Olate, J. (1994). Human-*Xenopus* chimeras of $G_s\alpha$ reveal a new

region important for its activation of adenylyl cyclase. *FEBS Letters*, **340**, 249–254.

Bax, A., Ikura, M., Kay, L. E. & Zhu, G. (1991). Removal of F1 baseline distortion and optimization of folding in multidimensional NMR spectra. *J. Magn. Reson.* **91**, 174–178.

Benjamin, D. R., Markby, D. W., Bourne, H. R. & Kuntz, I. D. (1995). Complete $^1H$, $^{13}C$ and $^{15}N$ assignments and secondary structure of the GTPase activating domain of $G_s$. *Biochemistry*, **34(1)**, 155–162.

Berlot, C. H. & Bourne, H. R. (1992). Identification of effector-activating residues of $G_{s\alpha}$. *Cell*, **68**, 911–922.

Birnbaumer, L. (1990). G proteins in signal transduction. *Annu. Rev. Pharmacol. Toxicol.* **30**, 675–705.

Bourne, H. R., Sanders, D. A. & McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117–127.

Brunger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Bruschweiler, R. & Case, D. A. (1994). Characterization of biomolecular structure and dynamics by NMR cross relaxation. *Prog. N.M.R. Spectrosc.* **26**, 27–58.

Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. (1990). Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell*, **62**, 317–329.

Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. & Sprang, S. (1994). Structures of active conformations of $G_{i\alpha 1}$ and the mechanism of GTP hydrolysis. *Science*, **265**, 1405–1412.

Conklin, B. R. & Bourne, H. R. (1993). Structural elements of $G\alpha$ subunits that interact with $G\beta\gamma$, receptors, and effectors. *Cell*, **73**, 631–641.

Davis, J. H. (1995). Refocusing revisited: an optimized, gradient-enhanced refocused HSQC and its applications in 2D and 3D NMR and in deuterium exchange experiments. *J. Biomol. NMR*, **5(4)**, 433–437.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graphics*, **6**, 13–27.

Freissmuth, M. & Gilman, A. G. (1989). Mutations of $G_{s\alpha}$ designed to alter the reactivity of the protein with bacterial toxins. Substitutions at $ARG^{187}$ result in loss of GTPase activity. *J. Biol. Chem.* **264(36)**, 21907–21914.

Fushman, D., Cahill, S., Lemmon, M. A., Schlessinger, J. & Cowburn, D. (1995). Solution structure of pleckstrin homology domain of dynamin by heteronuclear NMR spectroscopy. *Proc. Natl Acad. Sci. USA*, **92**, 816–820.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M. & Clore, G. M. (1994). The impact of direct refinement against three-bond HN-C$\alpha$ coupling constants on protein structure determination by NMR. *J. Magn. Reson. B*, **104(1)**, 99–103.

Grzesiek, S. & Bax, A. (1993). The importance of not saturating $H_2O$ in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594.

Grzesiek, S., Ikura, M., Clore, G. M., Gronenborn, A. M. & Bax, A. (1993). A 3D triple-resonance NMR technique for qualitative measurement of carbonyl-H$\beta$ *J* couplings in isotopically enriched proteins. *J. Magn. Reson.* **96**, 215–221.

Güntert, P. & Wüthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant angle constraints. *J. Biomol. NMR*, **1**, 447–456.

Güntert, P., Braun, W. & Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **317**, 517–530.

Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1(6)**, 736–751.

Iiri, T., Herzmark, P., Nakamoto, J. M., Dop, C. V. & Bourne, H. R. (1994). Rapid GDP release from $G_{s\alpha}$ in patients with gain and loss of endocrine function. *Nature*, **371**, 164–168.

Jones, D. T., Masters, S. B., Bourne, H. R. & Reed, R. R. (1990). Biochemical characterization of three stimulatory GTP-binding proteins: the large and small forms of $G_s$ and the olfactory specific G-protein, $G_{olf}$. *J. Biol. Chem.* **265**, 2671–2676.

Jurnak, F. (1985). Structure of the GDP domain of EF-Tu and location of the amino acids homologous to *ras* oncogene proteins. *Science*, **230**, 32–36.

Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665.

Kehlenbach, R. H., Matthey, J. & Huttner, W. B. (1994). XL$\alpha$s is a new type of G protein. *Nature*, **372(6508)**, 804–809.

Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. (1994). The 1.8 Å crystal structure of transducin $\alpha$-GDP: structural determinants for activation of a heterotrimeric G-protein $\alpha$ subunit. *Nature*, **369(6482)**, 621–628.

Landis, C. A., Masters, S. B., Spada, A., Pace, A. M., Bourne, H. R. & Vallar, L. (1989). GTPase inhibiting mutations activate the $\alpha$ chain of $G_s$ and stimulate adenylyl cyclase in human pituitary tumours. *Nature*, **340**, 692–696.

Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.

Markby, D. W., Onrust, R. & Bourne, H. R. (1993). Separate GTP binding and GTPase activating domains of a $G\alpha$ subunit. *Science*, **262**, 1895–1901.

Martin, J. R., Craven, C. J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V. & Waltho, J. P. (1995). The three-dimensional solution structure of human stefin A. *J. Mol. Biol.* **246**, 331–343.

Masters, S. B., Stroud, R. M. & Bourne, H. R. (1986). Family of G protein $\alpha$ chains: amphipathic analysis and predicted structure of functional domains. *Protein Eng.* **1**, 47–54.

Millhauser, G. L. (1995). Views of helical peptides–a proposal for the position of $3_{10}$-helix along the thermodynamic folding pathway. *Biochemistry*, **34(12)**, 3873–3877.

Moore, J. M., Lepre, C., Gippert, G. P., Chazin, W. J., Case, D. A. & Wright, P. E. (1991). High-resolution solution structure of reduced French bean plastocyanin and comparison with the crystal structure of poplar plastocyanin. *J. Mol. Biol.* **221**, 533–555.

Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. & Kay, L. E. (1993). A gradient $^{13}C$ Noesy-HSQC experiment for recording NOESY spectra of $^{13}C$-labeled proteins dissolved in $H_2O$. *J. Magn. Reson. B*, **102**, 317–321.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Noel, J. P., Hamm, H. E. & Sigler, P. B. (1993). The 2.2 Å crystal structure of transducin α-GTPγs. *Nature*, **366**, 654–663.

Quan, F., Thomas, L. & Forte, M. (1991). *Drosophila* stimulatory G protein α subunit activates mammalian adenylyl cyclase but interacts poorly with mammalian receptors: Implications for receptor-G protein interaction. *Proc. Natl Acad. Sci. USA*, **88**, 1898–1902.

Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. (1992). The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell*, **70**, 401–410.

Sigal, I. S., Gibbs, J. B., D'Alonzo, J. S. & Scolnick, E. M. (1986). Identification of effector residues and a neutralizing epitope of Ha *ras* p21. *Proc. Natl Acad. Sci. USA*, **83**, 4725–4729.

Vuister, G. W. & Bax, A. (1993). Quantitative *J* correlation: a new approach for measuring homonuclear three-bond $J(H^NH^\alpha)$ coupling constants in $^{15}$N-enriched proteins. *J. Am. Chem. Soc.* **115(17)**, 7772–7777.

Wagner, G. (1993). Prospects for NMR of large proteins. *J. Biomol. NMR*, **3(4)**, 375–385.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.

Wilson, P. T. & Bourne, H. R. (1995). Fatty acylation of $\alpha_z$. *J. Biol. Chem.* **270(16)**, 9667–9675.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley, New York.

Zhu, G. & Bax, A. (1992). Improved linear prediction of damped NMR signals using modified ''forward-backward'' linear prediction. *J. Magn. Reson.* **100**, 202–207.

***Edited by F. E. Cohen***

*(Received 12 July 1995; accepted 12 September 1995)*